

corporatings

AI performance on sustainability and accounting *without and with XBRL*

mhoullier@corporatings.com

On the performance of *raw* AI

On the performance of *raw* AI

Assessing Large Language Models Used for Extracting Table Information from Annual Financial Reports

by David Balsiger ¹, Hans-Rudolf Dimmler ¹ , Samuel Egger-Horstmann ¹ and Thomas Hanne ^{2,*}  

Published: 9 October 2024

Table 3. Combined statistics for BARD and ChatGPT-4.

Combined Statistics	Full Correct	Semi-Correct	Incorrect	% Full Correct	% Semi-Correct	% Incorrect
Total answers	501	57	130	72.80%	8.30%	18.90%
Total answers excluding						
currency	331	57	128	64.10%	11.00%	24.80%
Inventories	67	9	10	77.90%	10.50%	11.60%
Assets	66	14	6	76.70%	16.30%	7.00%
Liabilities	60	13	13	69.80%	15.10%	15.10%
Equity	64	13	9	74.40%	15.10%	10.50%
EBIT	34	4	48	39.50%	4.70%	55.80%
Gross profit	40	4	42	46.50%	4.70%	48.80%
Currency Balance sheet	85	0	1	98.80%	0.00%	1.20%
Currency Income statement	85	0	1	98.80%	0.00%	1.20%

AI extraction from unstructured documents

Sustainability Reporting Navigator

CSRD Reports Practices Requirements Preferences Resources More

	Company	Download	Country	Sector	Industry	Published	Pages	Auditor
	Verbund AG*	Link	Austria	Infrastructure	Electric Utilities & Power Generators	14.03.2025	192	EY
<input checked="" type="checkbox"/>	Vestas	Link	Denmark	Renewable Resources & Alternative	Wind Technology & Project Develop	05.02.2025	82	Deloitte
	VINCI	Link	France	Infrastructure	Engineering & Construction Services	28.02.2025	91	PwC
	Vivendi*	Link	France	Services	Media & Entertainment	07.04.2025	92	Deloitte & G
	Volkswagen	Link	Germany	Transportation	Automobiles	11.03.2025	240	EY
	Vonovia SE	Link	Germany	Infrastructure	Real Estate	19.03.2025	108	PwC
	VP Bank AG*	Link	Liechtenstein	Financials	Asset Management & Custody Activi	13.03.2025	106	PwC
	Wienerberger*	Link	Austria	Extractives & Mineral Processing	Construction Materials	31.03.2025	110	Deloitte
	Wirtualna Polska Holding*	Link	Poland	Technology & Communications	Internet Media & Services	02.04.2025	87	PwC
	WithSecure	Link	Finland	Technology & Communications	Software & IT Services	12.02.2025	76	PwC
	Wolters Kluwer	Link	Netherlands	Services	Professional & Commercial Services	12.03.2025	64	Deloitte

Search Engine

Reports marked with an asterisk (*) cannot yet be queried. We will upload them soon!

Search in the report of Vestas

- <https://sustainabilityreportingnavigator.com/#/csrdreports> (Mistral-large 24.11)
- ChatGPT 4o
- Gemini 2 Pro

AI extraction from unstructured documents

E3-4

Water consumption

1,000 m ³	2024
Total water consumption	73
Water consumption in areas at water risk	63
Water recycled and reused	0
Water stored	3
Changes in water storage	0
Total water withdrawals	323
Total water discharges	250
Water intensity (1,000 m ³ consumption per mEUR revenue)	0.004
Share of the measure obtained from direct measurement, from sampling etc. (%)	99

Vestas



What was the company's water intensity ratio in 2024?



In 2024, the company's water intensity ratio was 0.4 thousand m³ per mEUR revenue.

Scale error – actual: 0.4 m³ per million EUR

AI extraction from unstructured documents

Biological materials and reused or recycled materials used for manufacturing products •

	2024
Percentage of biological materials used in manufacturing (%)	12.1%
Reused or recycled materials used in manufacturing and packaging (tonnes)	50
Reused or recycled materials used in manufacturing and packaging (%)	1.62%

Demant



What was the percentage of biological materials (and biofuels used for non-energy purposes) used to manufacture the undertaking's products and services (including packaging) that is sustainably sourced in 2024?



The percentage of biological materials used in manufacturing that is sustainably sourced in 2024 is 12.1%.

Metric error – this is the share of all materials that are biological, instead of the share of biological materials that are sustainably sourced.

AI extraction from unstructured documents

ESRS ref.	Resource inflows, tonnes	2024
Technical materials for construction of new assets		
// E5-4, 31(a)	Steel	296,200
// E5-4, 31(a)	Copper	6,900
// E5-4, 31(a)	Aluminium	5,600
// E5-4, 31(a)	Plastics	9,700
// E5-4, 31(a)	Glass fibre	5,900
// E5-4, 31(a)	Rare earth elements	300
// E5-4, 31(a)	Concrete	6,800
// E5-4, 31(a)	Glass	26,600
Technical materials, scrap steel used in steel production		
// E5-4, 31(c)	In absolute value	59,200 - 103,700
// E5-4, 31(c)	In percentage, %	20 - 35

The use of scrap steel is a norm in steel production, with its content varying across geographies and reflecting established industry practices. Approximately 80% of the steel we source used in the production of steel

plates for foundations comes from Europe, where supplier data indicates that, on average, 35% of the material used in these plates derive from scrap.

Ørsted



What percentage of the company's input materials are recycled?



The percentage of the company's input materials that are recycled is approximately 35%.

Metric error – this is the share of steel that is ‘recycled’, instead of the share of all inputs that is recycled.

AI extraction from unstructured documents

Remuneration metrics

S1 – table 7

	2024
Gender pay gap ¹	-1,6%
CEO remuneration ratio	29

¹ Negative gender pay gap reflects a pay gap in favor of males

Bavarian Nordic



What was the unadjusted gender pay gap in 2024?



The unadjusted gender pay gap in 2024 was reported as -1.6%.

Sign error – the unadjusted gender pay gap as defined in the standard is +1.6%

AI extraction from unstructured documents



Airbus Annual Report 2024.pdf
PDF

What was the interest expense of Airbus during the reporting period?

Give the final answer as a JSON object, with a 'value' and a 'currency' attributes.

Explain where and how you found the value.



Airbus SE – IFRS Consolidated Income Statement for the years ended 31 December 2024 and 2023

(In € million)	Note	2024
Revenue	12	69,230
Cost of sales		(58,555)
Gross margin	12	10,675
Selling expenses		(877)
Administrative expenses		(1,744)
Research and development expenses	13	(3,250)
Other income	14	303
Other expenses	14	(197)
Share of profit from investments accounted for under the equity method	15	350
Other income from investments	15	44
Profit before financial result and income taxes		5,304
Interest income	16	802
Interest expense	16	(900)
Other financial result	16	219
Total financial result	16	121
Income taxes	17	(1,347)
Profit for the period		4,078

AI extraction from unstructured documents

13. Research and Development Expenses

Research and development expenses remained stable at € 3,250 million compared to € 3,257 million in 2023.

Research and development expenses mainly reflect the development of latest generation commercial aircraft programmes, including the A321XLR, A350 Freighter and activities to prepare technologies of the future.

The interest expense of Airbus during the reporting period was:

json

📄 Copier ✎ Modifier

```
{
  "value": 258,
  "currency": "EUR million"
}
```

Source and Explanation:

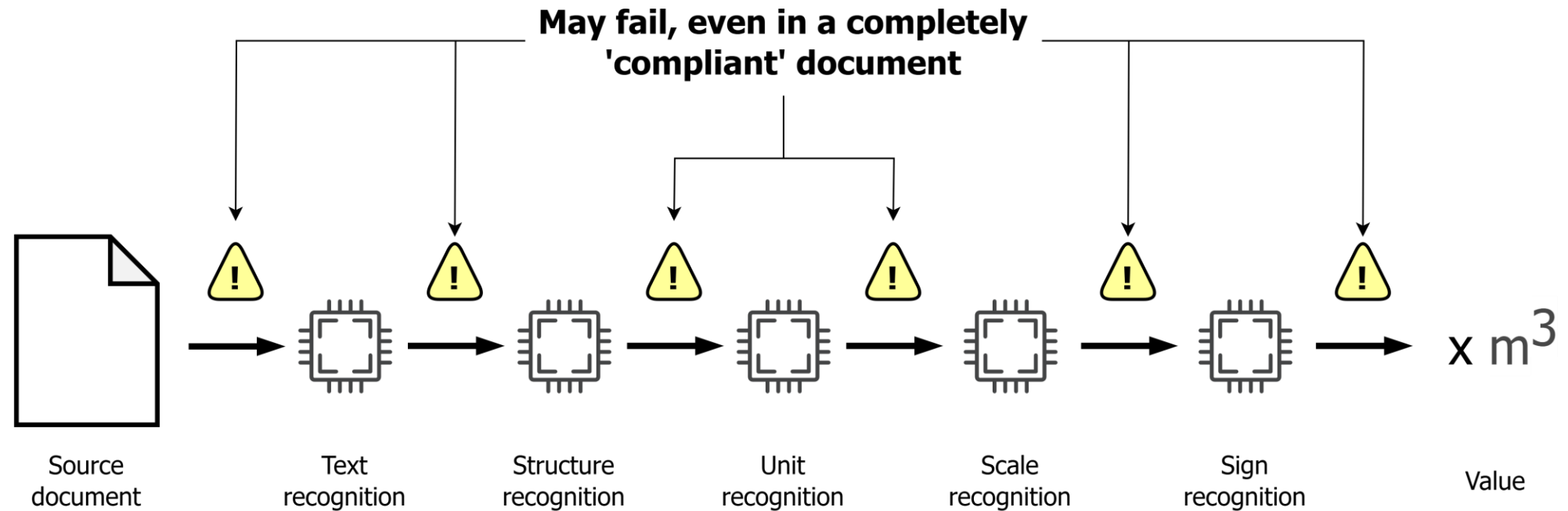
This value was found in the Airbus Annual Report 2024, under the section "Notes to the IFRS Consolidated Financial Statements — Note 13: Interest Result". The report states:

"Interest expense on financial liabilities measured at amortised cost amounted to €258 million for the year ended 31 December 2024".

Metric error – interest expense *on financial liabilities* instead of interest expense

Content error – the 'source' for the extracted value does not exist

AI extraction from unstructured documents



AI extraction from unstructured documents

What are [company]'s scope 1 emissions?



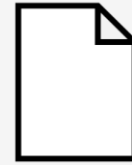
x1



x1

Approx. 5 seconds

What companies have reduced their scope 1 emissions the most in 2024?



x10000



x1

Approx. 14 hours

Database setup



x10000

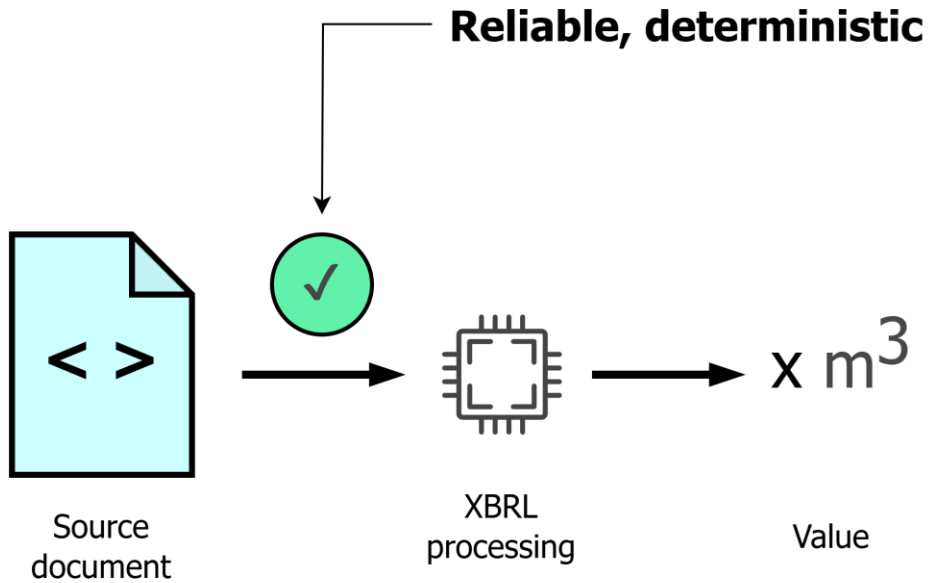


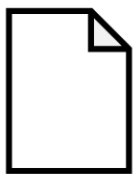
x1000

Approx. 580 days

This is parallelized in practice, but the overall **computing cost** remains the same.

Extraction from XBRL documents




x10000


x1

<1 second



Queries are technical

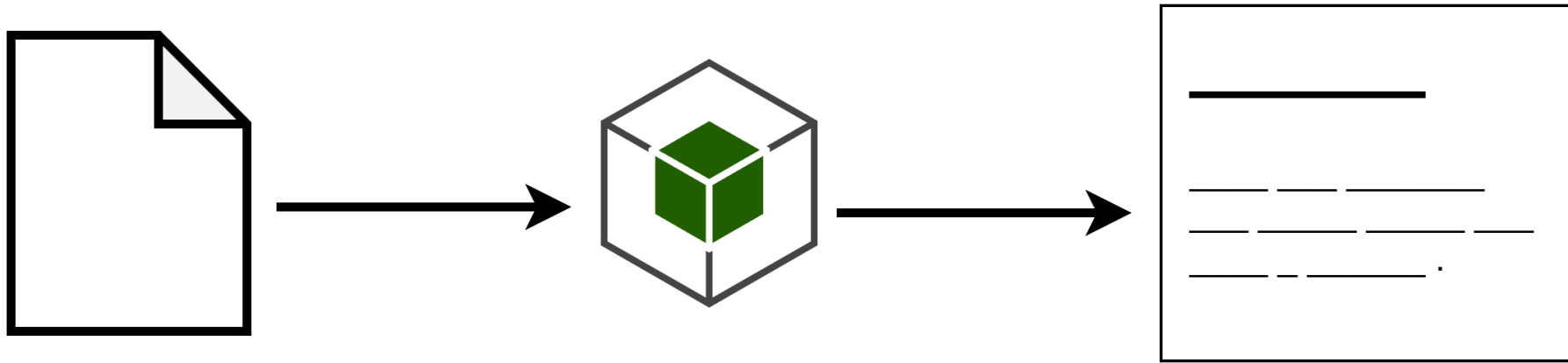


Only tagged information accessible



Only tagged documents accessible

Sources of AI performance



It is common to imagine that AI works like this: a user asks a question that is sent to a LLM, and the answer of the LLM is displayed to the user.

As a consequence, the increased performance of AI is often mistaken as an improvement of the LLMs.

Sources of AI performance

How much is $\exp(29)/(\ln(345))$?

The value of $\frac{e^{29}}{\ln(345)}$ is approximately 672,765,365,774.57. [↗]

This is **correct**.

How much is $\exp(29)$?

The value of $\exp(29)$, which is the same as e^{29} , is approximately:

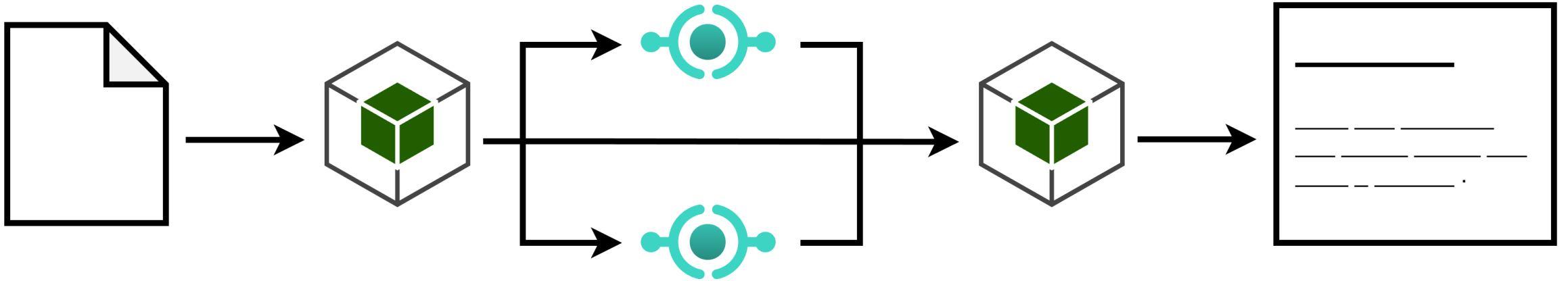
$$\exp(29) \approx 2.446 \times 10^{12}$$

That's about 2.45 trillion.

This is **incorrect** (actual value is +50%)

LLMs are not much better at mental calculations than a few years ago. The first example works because *the LLM is not the one doing the calculation*.

Sources of AI performance



Most of the progress of AI comes from improvements in its ability to **call other non-AI tools** (e.g., a calculator, a web search), and to rephrase their output.

Getting accurate answers from AI to accounting or sustainability question will only happen in the future if there are **non-AI tools designed to provide accurate answers to accounting or sustainability questions**.

Complementing AI with XBRL

Idea n°1: AI reading Inline XBRL



Idea n°1: AI reading Inline XBRL

AI determinants of success and failure: The case of financial statements

Marcelo Farr
Universidad Adolfo
Ibáñez

William C. Johnson
University of
Massachusetts

Ariel Markelevich*
Suffolk University

Alexis Montecinos
Suffolk University

* Corresponding author amarkelevich@suffolk.edu

March 2025

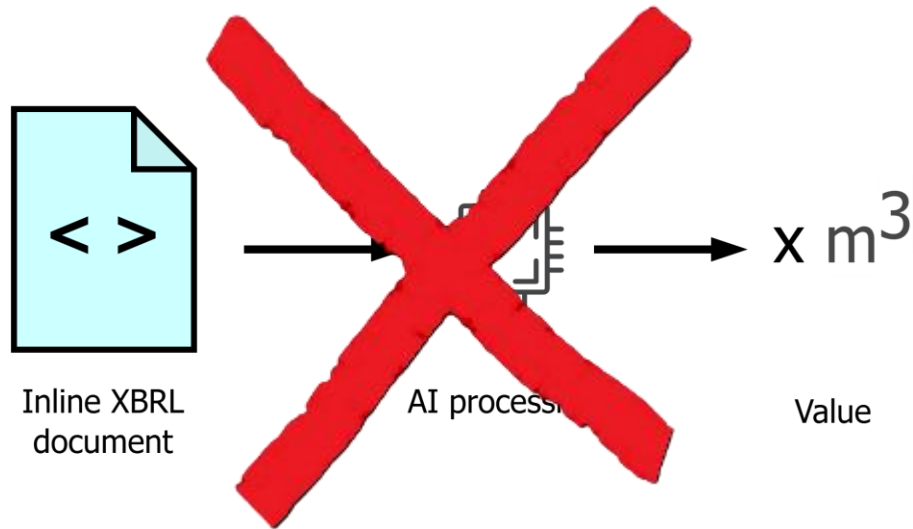
Table 3. AI Error rates for accounting data

	No context N=46,830	AI with text N=82,417	AI with HTML N=77,474	AI with XBRL N=87,978
Overall average across all concepts	99.91%	18.38%	16.33%	9.20%

LLMs do understand structured formats *significantly* better than unstructured ones.

Errors however still appear, and...

Idea n°1: AI reading Inline XBRL

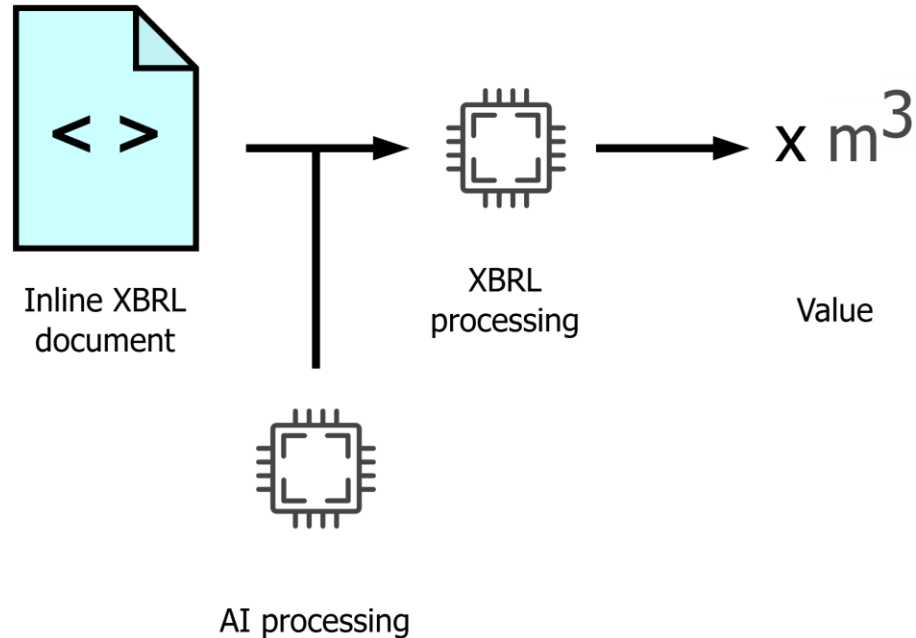


Speed performance of raw AI



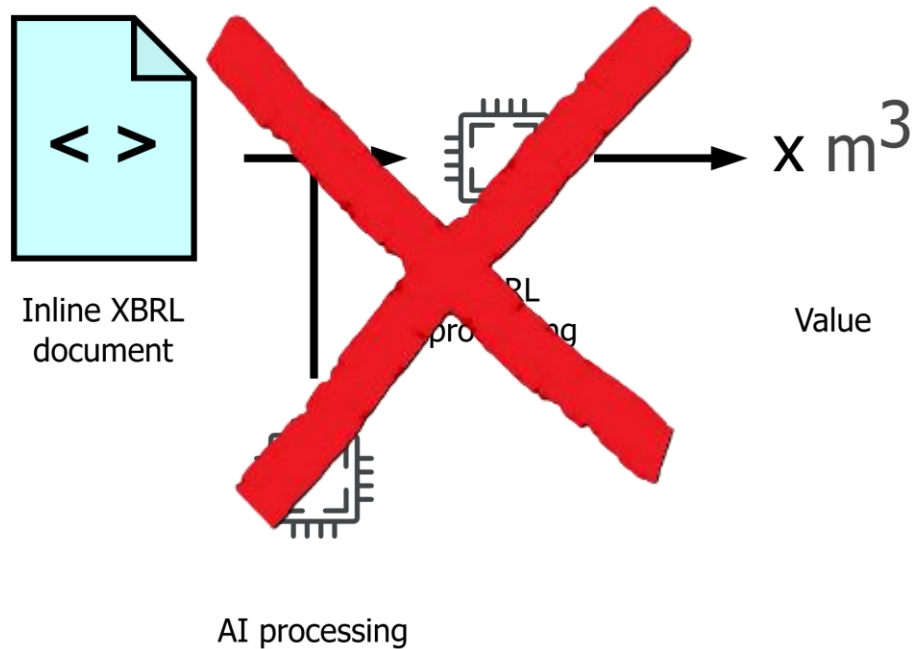
Relies on identifier names more than content

Idea n°2: AI handles technical parts of XBRL



- ✓ Better accuracy than AI on PDF
- ✓ Speed performance of XBRL
- ✓ Natural language queries

Idea n°2: AI handles technical parts of XBRL



⚠ Only tagged information accessible

⚠ Only tagged documents accessible

Idea n°3: Inline XBRL 'teaches' the AI



Is this a pyrophone?

Idea n°3: Inline XBRL 'teaches' the AI



Wendelin Weißheimer
playing the pyrophone

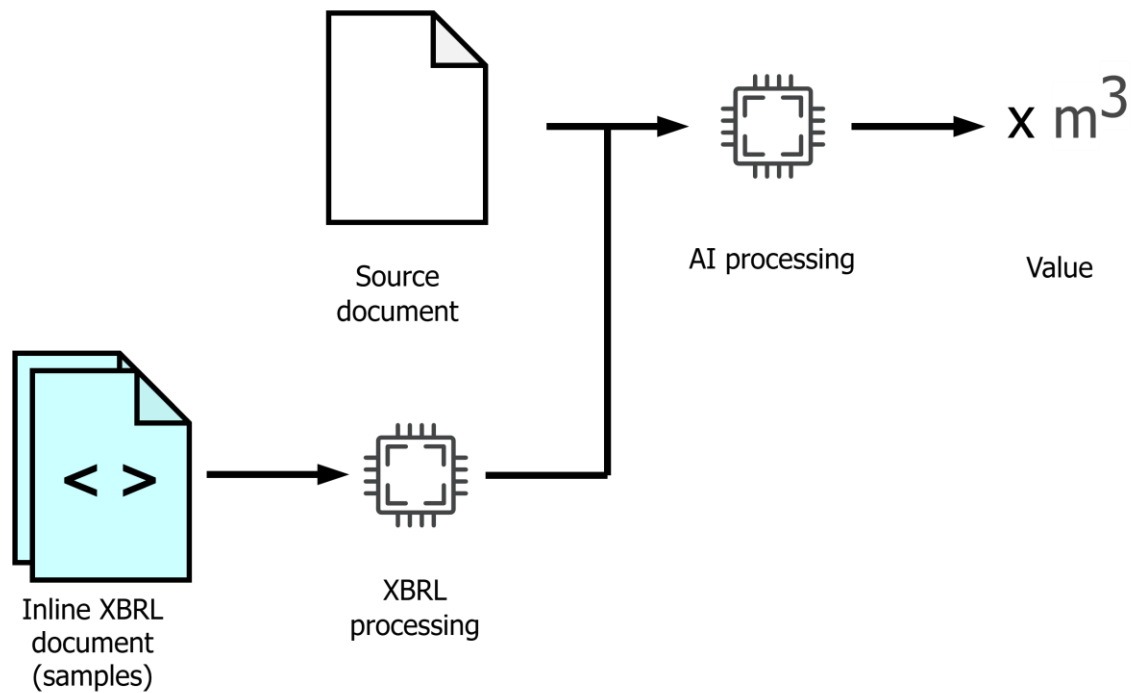


Is this a pyrophone?

Common pattern in modern AI
when too costly to *fine-tune* (re-
train).

Just as for humans, a single
example can be enough to unlock
the ability to answer a question.

Idea n°3: Inline XBRL 'teaches' the AI

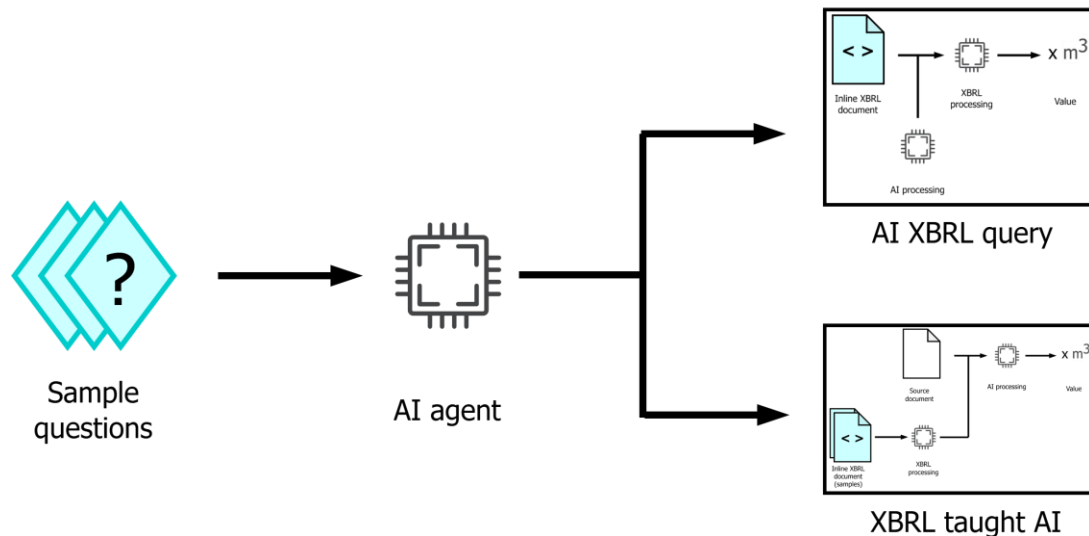


- ✓ Better accuracy than AI on PDF
- ✓ Natural language queries
- ✓ Not limited to tagged information
- ✓ Not limited to tagged documents



Speed performance of raw AI

Idea n°4: the agent



- ✓ Speed performance of XBRL
- ✓ Better accuracy than AI on PDF
- ✓ Natural language queries
- ✓ Not limited to tagged information
- ✓ Not limited to tagged documents

A LLM analyzes the user queries and decides to either:

- Create a technical XBRL query (Idea n°2, best speed and accuracy but only on XBRL documents)
- Create examples from XBRL (Idea n°3)

Idea n°4: the agent

Me

What was the interest expense of Airbus in 2024 ?

Corpo-Gen-AI

```
{
  issuer_name : 'AIRBUS SE',
  element_name : 'InterestExpense',
  value : 900000000,
  unit : 'EUR',
  period : 2024,
  source_link : {'source': 'CORPOCUBE', 'element_id': 3048, 'issuer_id': 13474, 'value': '900,000,000', 'unit': 'EUR', 'period': '2024', 'URL_link': 'http://ENDPOINT_NAME?externalLink=/report-viewer?reportId=487671&idInReport=auto-29'}
}
```

Obtained on the same Airbus **PDF** as earlier, **before the publication of Airbus' Inline XBRL document, using AI in combination with Airbus' and other companies' previous Inline XBRL publications.**

The AI/XBRL combination is also what enables our AI to learn *how* metrics are commonly referred to, and therefore accurately match a query to a figure.

Complementing AI with XBRL

- On raw PDF, AI shows unsatisfactory performance for value retrieval
 - main issues are reliability and time performance
 - even worse on narrative ('Does the company disclose restoration actions related to pollution?')
 - the lack of performance will not fix itself over time
- XBRL and AI can compensate each other's issues, if done properly
 - ❌ plugging AI into XBRL documents
 - ❌ using AI to handle the technical aspects of XBRL
 - ✅ using XBRL to feed AI with examples and counterexamples
 - ✅ using AI to decide between technical XBRL queries and further AI

AI performance with and without XBRL

- Thank you!
- Happy to further discuss: Marc Houllier,
mhoullier@corporatings.com
- Try AI on PDF yourself:
<https://sustainabilityreportingnavigator.com/#/csrdreports>